



## The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference

Trolle, Thomas; McMurtrey, Curtis P.; Sidney, John; Bardet, Wilfried; Osborn, Sean C.; Kaeffer, Thomas; Sette, Alessandro; Hildebrand, William H.; Nielsen, Morten; Peters, Bjoern

*Published in:*  
Journal of Immunology

*Link to article, DOI:*  
[10.4049/jimmunol.1501721](https://doi.org/10.4049/jimmunol.1501721)

*Publication date:*  
2016

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Trolle, T., McMurtrey, C. P., Sidney, J., Bardet, W., Osborn, S. C., Kaeffer, T., Sette, A., Hildebrand, W. H., Nielsen, M., & Peters, B. (2016). The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference. *Journal of Immunology*, 196(4), 1480-1487. <https://doi.org/10.4049/jimmunol.1501721>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Published in final edited form as:

*J Immunol.* 2016 February 15; 196(4): 1480–1487. doi:10.4049/jimmunol.1501721.

## The length distribution of class I restricted T cell epitopes is determined by both peptide supply and MHC allele specific binding preference

Thomas Trolle<sup>\*</sup>, Curtis P. McMurtrey<sup>†</sup>, John Sidney<sup>‡</sup>, Wilfried Bardet<sup>†</sup>, Sean C. Osborn<sup>†</sup>, Thomas Kaeffer<sup>‡</sup>, Alessandro Sette<sup>‡</sup>, William H. Hildebrand<sup>†</sup>, Morten Nielsen<sup>\*,§</sup>, and Bjoern Peters<sup>‡</sup>

<sup>\*</sup> Center for Biological Sequence Analysis, Department of Systems Biology, The Technical University of Denmark, Kongens Lyngby, Denmark

<sup>†</sup> Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA

<sup>‡</sup> Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, California, USA

<sup>§</sup> Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, San Martín, Buenos Aires, Argentina

### Abstract

HLA class I binding predictions are widely used to identify candidate peptide targets of human CD8<sup>+</sup> T cell responses. Many such approaches focus exclusively on a limited range of peptide lengths, typically 9 and sometimes 9-10 amino acids, despite multiple examples of dominant epitopes of other lengths. Here, we examined if epitope predictions can be improved by incorporating the natural length distribution of HLA class I ligands. We found that while different HLA alleles have diverse length binding preferences, the length profiles of ligands that are naturally presented by these alleles are much more homogeneous. We hypothesized that this is due to a defined length profile of peptides available for HLA binding in the endoplasmic reticulum. Based on this, we created a model of HLA allele specific ligand length profiles, and demonstrate how this model, in combination with HLA binding predictions, greatly improves comprehensive identification of CD8<sup>+</sup> T cell epitopes.

### Introduction

The identification of HLA class I (HLA-I) restricted epitopes recognized by human T cells has greatly benefited from the development of reliable binding prediction tools for different HLA molecules. For a given HLA molecule and a given peptide length, several benchmarks have shown that binding predictions correlate well with measured binding affinities (1–4), and that peptides with high predicted affinity contain the vast majority of T cell epitopes (5,

6). This has allowed comprehensive mapping of epitopes in entire pathogens by focusing testing on a manageable number of top predicted binders, saving vast amounts of resources (7–12).

However, it is not clear how peptides of different lengths should be treated in such prediction-guided approaches. Traditionally, there has been a focus on 9mer peptides when mapping HLA-I restricted T cell epitopes, but peptides of other lengths can bind HLA-I molecules (13) and elicit immune responses as evidenced by multiple dominant epitopes of length 8, 10 and 11 (14–17), and occasionally much longer peptides up to length 15 (17–19). MHC binding predictions for peptides of non-canonical lengths are available, but in many cases their predictions are extrapolated from 9mer data (20) and will predict a roughly similar affinity range for peptides of any given length. Thus, when considering all peptides of length 8–15 that have predicted affinities stronger than a given threshold, the number of peptide candidates would go up drastically compared to when only 9mers are considered.

The length distribution of T cell epitopes should largely reflect the length distribution of peptide ligands that are presented to T cells by MHC molecules. In turn, the MHC ligand length distribution should reflect at least two factors: The MHC allele specific ability to bind peptides of different lengths, and the MHC allele independent availability of peptides of different lengths for binding to MHCs, which is shaped by the antigen processing and presentation machinery preceding MHC binding, such as proteasomal cleavage and TAP transport (21). The goal of this study was to determine what the length distribution of MHC class I (MHC-I) restricted ligands is, to what degree this length distribution is allele specific, and how this knowledge can be utilized to optimize MHC-I binding predictions for CD8+ T cell epitope mapping.

## Materials and Methods

### MHC binding assays

Performance of quantitative *in vitro* competitive binding assays utilizing purified MHC-I and an iodine<sup>125</sup>-labeled standard probe peptide were performed using a monoclonal antibody capture assay platform essentially as described previously (22). Briefly, 0.1–1 nM of radiolabeled peptide was co-incubated at room temperature with 1  $\mu$ M to 1 nM of purified MHC-I in the presence of a cocktail of protease inhibitors and 1  $\mu$ M human  $\beta$ -2-microglobulin (Scripps Laboratories). Following a two-day incubation, MHC-I bound radioactivity was determined by capturing MHC-I/peptide complexes on W6/32 (anti-HLA class I monoclonal antibody)-coated Lumitrac 600 plates (Greiner Bio-one, Frickenhausen, Germany), and measuring bound radioactivity using the TopCount (Packard Instrument Co., Meriden, CT) microscintillation counter. The concentration of peptide yielding 50% inhibition of the binding of the radiolabeled peptide was calculated. Under the conditions utilized, where [label] < [MHC] and  $IC_{50} \ll [MHC]$ , the measured  $IC_{50}$  values were reasonable approximations of the true  $K_D$  values (23, 24). Each competitor peptide was tested at six different concentrations covering a 100,000-fold dose range, and in three or more independent experiments. As a positive control, the unlabeled version of the radiolabeled probe was also tested in each experiment.

27 HLA alleles representative of the most frequent HLA-I specificities in the human population were considered (14). Their peptide length preferences were determined by testing panels of combinatorial peptide libraries. Each library contained peptides of a uniform fixed length. Libraries ranged from 8–15 amino acid residues. Furthermore, each library, for each length, was defined by a fixed C-terminal residue that was either I, K or F. For each HLA allele, the libraries with the C-terminal residue that gave the highest affinity were chosen to determine the length preference for that HLA allele, in order to take into account the different C-terminal binding preferences of different alleles. It should be noted that the length binding profile for a given HLA molecule is estimated from a series of independent peptide libraries of different length. Due to variations in library synthesis, and binding assay variability, substantial noise is hence to be expected in the measured binding values leading to some degree of non-monotonic behavior of the length profile curves.

### Cell lines and Production of HLA complexes for elution studies

HeLa cells were cultured and propagated in DMEM with 10% FCS. HeLa cells were stably transfected with a soluble form of HLA-A\*01:01, HLA-A\*02:01, HLA-A\*24:02, HLA-B\*07:02 and HLA-B\*51:01 as previously described (25, 26). Soluble HLA (sHLA) constructs were generated with a truncation at the trans-membrane and cytoplasmic domains with the addition of a VLDLr purification tag and cloned into pcDNA3.1. HeLa cells were stably transfected with the sHLA by electroporation followed by drug selection and sub-cloning. sHLA producing clones were identified using a capture ELISA with the pan-class I antibody W6/32 as the capture antibody and a  $\beta$ -2-microglobulin antibody as a detector. sHLA producing clones were expanded and seeded into a hollow fiber bioreactor where sHLA containing supernatant was collected. sHLA was purified from the supernatant using affinity chromatography with an anti-VLDLr antibody. Complexes were eluted from the column in 0.2 M acetic acid and immediately processed for isolation of the peptide ligands.

### Elution of naturally presented MHC class I ligands

Eluted MHC-I ligand datasets were generated for five common HLA alleles: HLA-A\*01:01, HLA-A\*02:01, HLA-A\*24:02, HLA-B\*07:02 and HLA-B\*51:01. The elution of the peptide ligands was done as described in detail previously (25, 26). Briefly, the peptide ligands were eluted from the complex with an acid boil and peptide ligands were separated from the  $\alpha$  and  $\beta$  chains with a 3kDa cut-off filtration using a Millipore 3kDa NMWL ultrafiltration membrane (Merck Millipore). A 3kDa cutoff corresponds to the cutoff of a peptide approximately 30 amino acids. Since we use a maximum of 15mers in our predictions (or 1.5 kDa which is half the NMWL of the filter) we should have little to no bias in the number of 15mer peptides. However, the filtration efficiency may not be identical for all peptides between 8-15 residues long, and longer peptides may be underrepresented. Peptide pools were initially separated in to approximately 40 fractions using pH10 RP HPLC. Peptide containing fractions (fractions 22-60) were then analyzed individually with LCMS. Nano LC was performed with an Eksigent nanoLC4000 with an Eksigent autosampler (AB Sciex). Fractions were loaded on a C18 trap (350  $\mu$ m (i.d.) by 0.5 mm long; ChromXP) and desalted before separated with a gradient elution into a ChromXP C18 separation column (75  $\mu$ m (i.d.) by 15 cm long; ChromXP). Column media consisted of 3  $\mu$ m particles with 120 Å pores. The elution mobile phase consisted of two linear gradients

using solvent A (98% water, 2% acetonitrile, 0.1% formic acid) and solvent B (95% acetonitrile, 5% water, 0.1% formic acid): 10% to 40% B for 70 minutes then 40% to 80% for 10 minutes. Eluate was ionized with a Nanospray III ion source (AB Sciex) and MS1 and MS2 fragment were obtained in IDA mode using a AB Sciex 5600 Triple TOF as described previously (26).

Peptide sequences were derived from the resulting fragment spectra using PEAKS 7.0 (Bioinformatic Solutions) with a precursor ion tolerance of 50 ppm and product ion tolerance of 0.05Da. The NCBI non-redundant database with *H. sapiens* taxonomy was used. Post-translational modifications consisting of N-terminal acetylation, deamidation of Asn and Gln, oxidation of Met, His, Trp, sodium adducts of Asp, Glu, C-terminus, and the pyroglutamate derivative of glutamic acid, were searched as variable modifications. Positive sequence assignments were determined at a 1% FDR using the decoy fusion approach (27). Most positive peptide identifications were within 25 ppm of theoretical mass. Any peptides from the sHLA construct (HLA  $\alpha$  chain and  $\beta$ -2-microglobulin), and a contaminating protein TERA were removed from the data, as these are likely not ligands. Peptides resulting from a D|P, D|A, and D|T cleavages were also removed as these are peptides likely created from acid hydrolysis of larger ligands.

### Corrected elution datasets

In addition to the eluted peptide dataset described above, two corrected datasets were created. The first corrected dataset was obtained by filtering out ligands that did not conform to the canonical MHC binding motif of the given allele in order to remove likely contaminants. Binding affinities for all eluted peptides were predicted using *NetMHCpan-2.8* (28, 29). In addition to binding affinities in nM, *NetMHCpan* also returns a percentage rank score for each peptide, indicating how strong a peptide's binding affinity is compared to a large pool of naturally occurring peptides. A rank score of 10% means a peptide falls within the top 10% strongest binders among the pool of naturally occurring peptides. The standard *NetMHCpan* rank score is based on predicted binding affinities of 9mer peptides only. Here, we extend this and calculate the rank score compared to pools of peptides matching the length of the query peptide. This was done to remove any artificial bias in the rank scores imposed by the use of the extrapolation model from 9mer data mentioned earlier (20). A rank score of 10% was used as the threshold for defining a binder. Note, that this is a very tolerant threshold, as earlier studies have demonstrated that the vast majority of known CD8+ epitopes are predicted to bind to the restricting MHC molecules with a rank score less than or equal to 2% (5, 30).

The second and final corrected dataset took into account that some peptides are degraded before mass spectrometry identification, leading to the recovery of fragments of the original full-length ligand. To identify such peptide degradation events, we mapped all predicted non-binders back to their source proteins and extended the peptides *in silico* with up to five amino acids at either the N- or C-terminus, up to a maximum of length 15, while searching for potential predicted binders. If a high affinity binder, defined using a rank score threshold of 2%, was discovered in this process, it was substituted for the non-binding fragment. We decided to use a more stringent rank score threshold of 2% in this case, as we were only

interested in including extended peptides that had a very high likelihood of binding their given HLAs. In contrast, the previous filtering used a 10% rank score threshold, as there our goal was to only exclude peptides that had a very low probability of binding their given HLAs. It should be noted that similar results were obtained using an unfiltered data set, as well as with data sets where the thresholds for identifying peptide degradation events were 1% and 10% (data not shown).

### Reconstruction of the peptide length profile available for binding to MHC

Assuming that peptides available for binding to MHC (AP) can be approximated by a Boltzmann distribution, the ratio of the number of peptides of a given length  $L$  bound to MHC,  $P_{\text{MHC}}(L)$ , compared to the number of peptides bound of length 9,  $P_{\text{MHC}}(9)$ , is determined by the ratio of peptides available for binding of length  $L$ ,  $AP(L)$ , and those of length 9,  $AP(9)$ , and the difference in binding free energy of these peptides. In our assay conditions,  $\log(\text{IC}_{50})$  approximates binding free energy, and thus we can write:

$$\frac{P_{\text{MHC}}(L)}{P_{\text{MHC}}(9)} = \frac{AP(L)}{AP(9)} \cdot e^{-\beta \log\left(\frac{\text{IC}_{50}(L)}{\text{IC}_{50}(9)}\right)} = \frac{AP(L)}{AP(9)} \cdot \left(\frac{\text{IC}_{50}(9)}{\text{IC}_{50}(L)}\right)^{\beta}$$

where  $\beta$  is a positive unknown parameter, and the  $\text{IC}_{50}$  values and bound length distributions ( $P_{\text{MHC}}$ ) are known to us based on the affinity measurements and elution experiments for five HLA alleles. Thus, we can fit  $\beta$  and the unknown available peptide length distribution by minimizing the squared distance between measured and calculated  $P_{\text{MHC}}(L) / P_{\text{MHC}}(9)$  values.

### Benchmark data

A T cell epitope evaluation dataset was retrieved from the Immune Epitope Database (IEDB)(31). As it is of particular importance for our study that the optimal length peptide epitope was identified, we restricted ourselves to multimer/tetramer assays. Peptides between the lengths of 8-15 amino acids were included in which the tetramer utilized was one of the 27 IEDB reference HLA alleles. Source proteins for each epitope were downloaded from GenBank using the accession number annotated in the IEDB. A total of 535 T cell epitopes matching our selection criteria were downloaded. These epitopes were filtered to remove predicted non-binders using the same approach as for the elution dataset, reducing the dataset by 42 epitopes. Finally, five epitopes were removed from the dataset, as they could not be mapped to their annotated source protein. As a majority of the epitopes (59%) in the dataset were HLA-A\*02:01 restricted, we created a balanced dataset, where 20 epitopes for each HLA allele were selected at random. If there were less than 20 epitopes for an allele, all the epitopes were included in the balanced dataset. Binding affinity predictions were generated for each T cell epitope as well as all overlapping 8-13mers in the source proteins using *NetMHCpan*. As no 14-15mers were present in the final datasets, these lengths were excluded from the benchmark.

In addition to the T cell epitope dataset, three recently published MHC-I ligand datasets by Granados *et al.* (32), Marcilla *et al.* (33) and Thommen *et al.* (34) were retrieved from the

IEDB (IEDB reference IDs 1027559, 1027269 and 1027076). The datasets were filtered to remove predicted non-binders as previously described, removing 48, 81 and 27 peptides respectively. Binding affinity predictions were generated for overlapping peptides in the source proteins as described for the IEDB dataset. Overlapping 8-11mers were included in the Granados benchmark, while 8-13mers were included in the Marcilla and Thommen benchmarks, reflecting the ligand lengths found in each dataset.

### Adjusting binding affinity predictions for length preference

*NetMHCpan* predictions were adjusted to result in a distribution of predicted binders that reflect the length profile of ligands for a given HLA allele. This was achieved by dividing the predicted rank scores for each peptide of length  $L$  by the relative frequency at which peptides of this length are found bound to the MHC:

Adjusted peptide rank score = peptide rank score / ( $P_{\text{MHC}(L)} / P_{\text{MHC}(9)}$ ) where the ratio ( $P_{\text{MHC}(L)} / P_{\text{MHC}(9)}$ ) was estimated in an MHC specific manner using the model described above. This meant that 9mer predictions were left unchanged whereas predictions for all other lengths were modified depending on the peptide length and MHC restriction. Peptide lengths that were enriched compared to 9mers received enhanced length corrected rank scores, and vice versa for peptide lengths that were less preferred compared to 9mers for the given MHC. For peptides with MHC restrictions outside our panel of 27 HLA-I alleles, an averaged MHC binding length preference was used instead. This averaged length distribution (found in Supplemental Table I) was determined by calculating the geometric mean of the 27 HLA-I length preferences. Finally, the corrected rank scores were transformed back to binding affinity values using the underlying percentile affinity distribution for the given allele, which corresponds to an effective  $\text{IC}_{50}$  value.

## Results

### MHC binding length preference

We set out to determine the preferred length of peptides binding to a panel of 27 human MHC-I alleles comprising commonly expressed molecules in the human population. Affinities of combinatorial libraries of peptides with different lengths were determined and normalized to the affinity of a library of 9mer peptides as described in the Materials and Methods section. The resulting length preferences are shown in Supplemental Table I, and Fig. 1 depicts data for five alleles that are representative of the spectrum of observed patterns: HLA-A\*02:01, HLA-A\*24:02 and HLA-B\*07:02 showed the typical preference for 9mer peptides, while HLA-A\*01:01 had a preference for 10mers and HLA-B\*51:01 had a preference for 8mers. These data confirm that there are MHC allele specific differences in binding affinity for peptides of different length.

### Length distribution of naturally presented MHC ligands

To examine if allele specific differences in binding length preferences have an impact on the length distribution of naturally processed MHC ligands, we performed peptide elution studies on the five alleles for which binding data are shown in Fig. 1. Elution of peptide ligands from secreted MHCs and ligand identification by mass spectrometry was performed



as described previously (25, 26) and in the Materials and Methods section. To remove contaminants and to control for the effect of peptide degradation, the dataset was further corrected with the help of binding predictions. The final datasets, which can be found in Supplemental Table II, contain an average of 3,197 identified peptides per allele ranging from 1,275 for HLA-B\*51:01 to 4,456 for HLA-A\*02:01 as listed in Table I. These datasets have been submitted to the IEDB and provide a large publicly available dataset of naturally presented MHC ligands with well-defined restrictions for different alleles gathered with a consistent methodology. The data can be accessed at IEDB Submission ID <http://www.iedb.org/subID/1000685>.

Next, we examined the length distribution of the ligands identified in the elution studies. Fig. 2 shows, for each allele, the number of ligands identified for a given length normalized by the number of peptides identified for the HLA at length 9. Raw peptide counts and 9mer normalized values can be found in Supplemental Table II. Strikingly, all the HLAs presented more 9mers than peptides of any other lengths. This observation includes HLA-A\*01:01 and HLA-B\*51:01, which preferred binding 10mers and 8mers over 9mers, respectively. However, compared to HLA-A\*02:01, HLA-A\*24:02 and HLA-B\*07:02, all of which prefer binding of 9mers, HLA-A\*01:01 presented the most 10mer ligands, and HLA-B\*51:01 presented the most 8mers. This suggested that the HLA allele dependent length preferences observed in the binding assay did impact the length distribution of naturally presented ligands in an allele specific fashion, but that other factors dampened the allele specific effects and led to a predominant presentation of 9mer peptides.

### Length profiles of peptides available for MHC binding

We wanted to test the hypothesis that the discrepancy between allele specific peptide length binding preferences and naturally presented ligand repertoires was due to a fixed length distribution of peptides available for binding to MHCs. To do this, we built a simple mathematical model that assumed that the available peptide length distribution was the same for each MHC allele and that the observed eluted MHC ligand length profiles displayed in Fig. 2 were the result of this available peptide length profile in conjunction with the MHC-I binding length preferences of each allele. Based on this, we calculated the available peptide length profile shown in Fig. 3 and Table II. By far the most frequent peptide length available for binding based on this model was 9, which was expected given that this peptide length dominated in the eluted ligand profile for all alleles, even for HLA-A\*01:01 and HLA-B\*51:01 that preferred to bind peptides of different lengths.

To evaluate if the available peptide length profile we calculated based on data from five alleles was robust, a leave-one-out experiment was carried out. Iteratively, each HLA allele was excluded from the training data, an available peptide length profile was calculated using the remaining alleles, and the measured and predicted eluted peptide length profiles for the excluded HLA alleles were compared. Comparisons of the predicted and measured MHC ligand length profiles are shown in Fig. 4. For all alleles, there was an excellent fit of predicted and measured profiles (average RMSD =  $0.057 \pm 0.021$ ). This demonstrates that we are able to estimate one common available peptide length profile, which, when combined



with allele specific binding preferences, is able to explain the differences between our five naturally processed MHC ligand length profiles.

### Benchmarking length profile adjusted MHC binding predictions

Next we tested if our modeled peptide length distributions of naturally presented peptides could be utilized to improve the prediction of T cell epitopes and naturally processed MHC-I ligands. Rather than predicting candidate peptides based on binding affinity alone, we added a correction factor resulting in a length adjusted binding prediction for each peptide (see Materials and Methods). As shown in Supplemental Fig. 1, choosing peptides based on length-adjusted predictions resulted in a length distribution of peptides that mimicked that of naturally eluted ligands.

We evaluated the performance of the length adjusted binding affinity predictions on one T cell epitope dataset and three MHC-I ligand datasets. The T cell epitope dataset was retrieved by querying the IEDB for peptides that were recognized by human T cells in tetramer staining assays, which we considered most reliable to identify exact epitopes. This dataset was further balanced to not over-represent commonly studied alleles such as HLA-A\*02:01. The resulting dataset contained 185 epitopes ranging in length from 8–13 residues (Supplemental Table III). These epitopes were considered positives while all other 8-13mers from the same proteins were considered negatives. Peptides were pooled and sorted by predicted binding affinity (from strongest to weakest). This sorted peptide list was then used to determine the number of epitopes identified vs. the number of peptides tested. Plots were created using three different approaches for predicting T cell epitopes: 1) pure MHC binding predictions considering peptides of length 8-13 equally, 2) MHC binding predictions for 9mer peptides only and 3) the newly developed length-adjusted MHC binding predictions. Our goal here was to compare our novel length correction approach with two other prediction strategies that are currently utilized: predicting epitopes of multiple lengths and treating each length equally, and predicting epitopes for a single, optimal length. We opted to use 9 as the optimal length for all alleles as, for each allele in our study, this was the most common ligand length found. From the plots (Fig. 5, top left panel), it was apparent that the first approach of considering all peptide lengths equally had the worst performance, as, for example, approximately twice the number of peptides had to be considered to identify 60% of the epitopes in the benchmark. The other two approaches had very similar performances when considering the top 0.5% of peptides. However when the goal is to identify 80% or more of the epitopes, the length corrected approach is far superior to the 9mer only prediction approach, which will – by definition – miss epitopes of other lengths. Most importantly, this ability to comprehensively identify epitopes of all lengths comes with no significant additional cost, in contrast to the naive approach of considering all peptide lengths equally.

Next, we queried the IEDB for large scale MHC-I ligand elution datasets identified by different groups for which restrictions were determined and that represented peptides of different length. Three datasets were selected namely the “Granados MHC-I ligands” (4433 ligands, lengths 8-11)(32), the “Marcilla MHC-I ligands” (2235 ligands, lengths 8-13)(33) and the “Thommen MHC-I ligands” (1041 ligands, lengths 8-13)(34). Note, that all ligand

counts are after filtering for predicted non-binders. Eluted MHC ligands were considered positives and all other peptides from the same proteins were considered negatives. The results of these three benchmarks (shown in Fig. 5, top right and bottom panels) were very similar to the IEDB dataset benchmark. Considering peptides of all lengths equally was the worst approach, while both the 9mer only approach and the weighted length approach performed equally well when the goal was to discover a subset of the eluted ligands (<50-60%). But when the goal was to comprehensively predict eluted ligands (>60%), the length weighted approach was far superior to the other two. Thus, the length-adjusted prediction approach developed here performed as well or better than the two other approaches in three independent benchmarks, and was the most efficient approach for the comprehensive discovery of both epitopes and eluted ligands.

## Discussion

MHC binding predictions have facilitated T cell epitope discovery by narrowing the search space to a manageable number of likely peptide candidates. When compared to approaches that do not use predictions, such as screening overlapping peptides, a downside of the prediction approach was that peptides of non-canonical lengths were missed (35). Naively, one could simply extend binding predictions to peptides of any length, and rank all peptides based on their predicted affinity. But, as demonstrated in our study, while peptides of non-canonical lengths might have similar or even better predicted binding affinities than the canonical 9mer peptides, they end up being underrepresented among the naturally presented ligands eluted from MHC molecules, and consequently are also less frequently found to be recognized by T cells. In this study, we explained these similarities between the length profiles of naturally presented peptides by fitting a common, underlying peptide length distribution. This common length distribution, which we call the “available peptide length profile”, could be combined with allele specific binding length preferences to yield allele specific MHC ligand length profiles. However, at this point, we can only speculate on what the major factors behind the available peptide length profile are, as well what their relative contributions are.

The available peptide length profile (Fig. 3), suggests that 9mer peptides are the most common peptides available for MHC binding, with 8mers, 10mers and longer peptides being far less frequent. We hypothesize that this is due to a combination of three antigen processing mechanisms, the peptide cleavage by the proteasome, transport into the ER by TAP and peptide trimming by ERAP. Peptide fragments generated by the proteasome are generally 4-7 amino acids long, with the frequency of fragments longer than that decreasing as length increases (36–38). We see a similar decrease in the available peptide length profile from length 10 and upward, which thus could be attributed to the proteasome. TAP has been shown to preferentially transport peptides 9-16 amino acids long (39, 40), which would explain the low frequency of 8mers in the available peptide length profile. Finally, peptides in the ER are trimmed by ERAP down to a minimum length of 9 amino acids (41), explaining the 9mer peak in the available peptide length profile. Thus, while we did not experimentally verify the available peptide length profile, our fitted profile is consistent with previous knowledge of antigen processing.

While it is generally accepted that MHC-I molecules bind peptides with lengths 8-11, longer peptide binders have also previously been observed. These long peptide ligands can bind in a variety of configurations. Structural studies show that the majority of long ligands bind using the P2 and C-terminal amino acids with a central bulge to accommodate the increased length of the peptide (42, 43). Thus far, this appears to be the primary mechanism by which long MHC-I ligands are bound. A second method for binding has been suggested in the literature whereby a portion of the peptide binds in the groove with a C-terminal or N-terminal extension (44, 45). Indeed, there is a single structure of a peptide binding in an extended configuration (46). While there are examples of peptide ligands binding with a C-terminal extension, it is unknown how frequently this occurs. In our eluted ligand dataset there is some evidence of C-terminal extended peptides. However, here, we have separated these putative extended ligands from the canonical binding peptides and these extended ligands will be investigated in detail in future studies.

We have here developed a simple yet effective approach to adjust the predicted binding affinity of a peptide based on its length and the corresponding availability for peptide binding to MHC. We show that this is much more effective in identifying epitopes compared to considering peptides of all lengths equally at any threshold. Also, our novel approach compares favorably to the approach of considering only 9mer peptides when the goal is to comprehensively identify epitopes. While in our benchmark of tetramer mapped epitopes this effect was most pronounced when the goal was to identify more than 80% of all epitopes, it was already apparent when considering more than 50% of all ligands. Given that the tetramer mapping dataset is biased in that researchers will preferably make 9mer peptide based tetramers, we expect that the estimate based on the elution datasets is more accurate. Thus, we would suggest that the length based weighting of MHC binding predictions introduced here should be applied to any study aimed at comprehensively identifying MHC-I restricted epitopes.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

For so much more than just her invaluable contributions to the performance of the peptide binding assays described herein, we dedicate this work to the memory of Carrie Moore (1982-2015).

This project has been funded in whole or in part with Federal funds from the National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272201200010C. MN is a researcher at the Argentinean national research council (CONICET).

## Abbreviations

<b>ER</b>	endoplasmic reticulum
<b>HLA-I</b>	HLA class I
<b>IEDB</b>	Immune Epitope Database

**MHC-I** MHC class I  
**sHLA** soluble HLA

## References

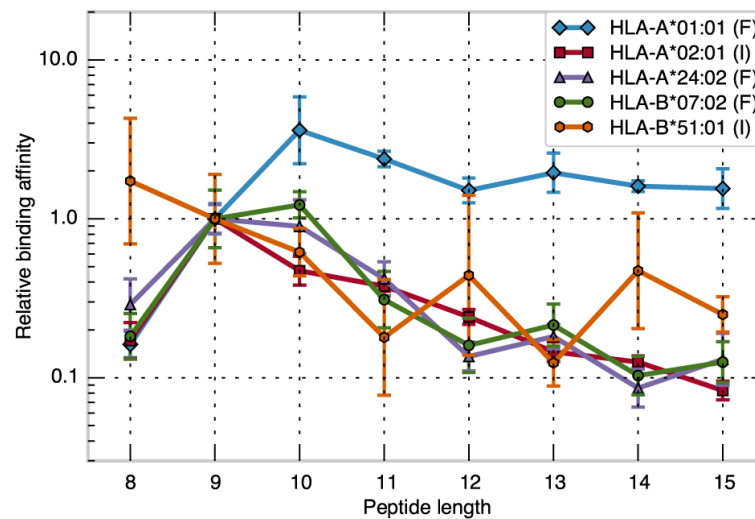
1. Peters B, Bui H-H, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson SS, Sidney J, Lund O, Buus S, Sette A. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput. Biol.* 2006; 2:e65. [PubMed: 16789818]
2. Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol.* 2008; 9:8. [PubMed: 18366636]
3. Zhang GL, Ansari HR, Bradley P, Cawley GC, Hertz T, Hu X, Jojic N, Kim Y, Kohlbacher O, Lund O, Lundegaard C, a Magaret C, Nielsen M, Papadopoulos H, Raghava GPS, Tal V-S, Xue LC, Yanover C, Zhu S, Rock MT, Crowe JE, Panayiotou C, Polycarpou MM, Duch W, Brusic V. Machine learning competition in immunology - Prediction of HLA class I binding peptides. *J. Immunol. Methods.* 2011; 374:1–4. [PubMed: 21986107]
4. Trolle T, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, Sette A, Peters B, Nielsen M. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics.* 2015; 31:2174–2181. [PubMed: 25717196]
5. Erup Larsen M, Kloverpris H, Stryhn A, Koofhethile CK, Sims S, Ndung'u T, Goulder P, Buus S, Nielsen M. HLArestrictor - a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides. *Immunogenetics.* 2011; 63:43–55. [PubMed: 21079948]
6. Paul S, Weiskopf D, a Angelo M, Sidney J, Peters B, Sette A. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J. Immunol.* 2013; 191:5831–9. [PubMed: 24190657]
7. Wang M, Lamberth K, Harndahl M, Røder G, Stryhn A, Larsen MV, Nielsen M, Lundegaard C, Tang ST, Dziegiel MH, Rosenkvist J, Pedersen AE, Buus S, Claesson MH, Lund O. CTL epitopes for influenza A including the H5N1 bird flu; genome-, pathogen-, and HLA-wide screening. *Vaccine.* 2007; 25:2823–2831. [PubMed: 17254671]
8. Sedegah M, Kim Y, Ganeshan H, Huang J, Belmonte M, Abot E, Banania JG, Farooq F, McGrath S, Peters B, Sette A, Soisson L, Diggs C, Doolan DL, Tamminga C, Villasante E, Hollingdale MR, Richie TL. Identification of minimal human MHC-restricted CD8+ T-cell epitopes within the *Plasmodium falciparum* circumsporozoite protein (CSP). *Malar. J.* 2013; 12:185. [PubMed: 23738590]
9. Chiu C, McCausland M, Sidney J, Duh FM, Roupheal N, Mehta A, Mulligan M, Carrington M, Wieland A, Sullivan NL, Weinberg A, Levin MJ, Pulendran B, Peters B, Sette A, Ahmed R. Broadly Reactive Human CD8 T Cells that Recognize an Epitope Conserved between VZV, HSV and EBV. *PLoS Pathog.* 2014:10.
10. Rajasagi M, Shukla S. a. Fritsch EF, Keskin DB, DeLuca D, Carmona E, Zhang W, Sougnez C, Cibulskis K, Sidney J, Stevenson K, Ritz J, Neuberger D, Brusic V, Gabriel S, Lander ES, Getz G, Hacohen N, Wu CJ. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood.* 2014; 124:453–462. [PubMed: 24891321]
11. Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, Lin JC, Teer JK, Clifton P, Tycksen E, Samuels Y, a Rosenberg S. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat. Med.* 2013; 19:747–52. [PubMed: 23644516]
12. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rehkman N, Moreira AL, Ibrahim F, Bruggeman C, Gasmi B, Zappasodi R, Maeda Y, Sander C, Garon EB, Merghoub T, Wolchok JD, Schumacher TN, Chan TA. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* 2015; 348:1–10. (80-. ).

13. Chen Y, Sidney J, Southwood S, Cox AL, Sakaguchi K, a Henderson R, Appella E, Hunt DF, Sette A, Engelhard VH. Naturally processed peptides longer than nine amino acid residues bind to the class I MHC molecule HLA-A2.1 with high affinity and in different conformations. *J. Immunol.* 1994; 152:2874–2881. [PubMed: 8144888]
14. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, Broadwater A, V Kolla R, De Silva AD, de Silva AM, Mattia KA, Doranz BJ, Grey HM, Shresta S, Peters B, Sette A. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:E2046–53. [PubMed: 23580623]
15. Lidehall AK, Sund F, Lundberg T, Eriksson BM, Tötterman TH, Korsgren O. T cell control of primary and latent cytomegalovirus infections in healthy subjects. *J. Clin. Immunol.* 2005; 25:473–481. [PubMed: 16160916]
16. Motozono C, Kuse N, Sun X, Rizkallah PJ, Fuller A, Oka S, Cole DK, Sewell AK, Takiguchi M. Molecular Basis of a Dominant T Cell Response to an HIV Reverse Transcriptase 8-mer Epitope Presented by the Protective Allele HLA-B\*51:01. *J. Immunol.* 2014; 192:3428–34. [PubMed: 24600035]
17. Rist MJ, Theodossis A, Croft NP, Neller M. a. Welland A, Chen Z, Sullivan LC, Burrows JM, Miles JJ, Brennan RM, Gras S, Khanna R, Brooks AG, McCluskey J, Purcell AW, Rossjohn J, Burrows SR. HLA peptide length preferences control CD8+ T cell responses. *J. Immunol.* 2013; 191:561–71. [PubMed: 23749632]
18. Tey SK, Goodrum F, Khanna R. CD8+ T-cell recognition of human cytomegalovirus latency-associated determinant pUL138. *J. Gen. Virol.* 2010; 91:2040–2048. [PubMed: 20375220]
19. Hassan C, Chabrol E, Jahn L, Kester MGD, de Ru AH, Drijfhout JW, Rossjohn J, Falkenburg JHF, Heemskerk MHM, Gras S, van Veelen P. a. Naturally Processed Non-canonical HLA-A\*02:01 Presented Peptides. *J. Biol. Chem.* 2015; 290:2593–2603. [PubMed: 25505266]
20. Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics.* 2008; 24:1397–8. [PubMed: 18413329]
21. Blum JS, a Wearsch P, Cresswell P. Pathways of Antigen Processing. *Annu. Rev. Immunol.* 2013; 31:443–473. [PubMed: 23298205]
22. Sidney, J.; Southwood, S.; Moore, C.; Oseroff, C.; Pinilla, C.; Grey, HM.; Sette, A. *Current Protocols in Immunology.* John Wiley & Sons, Inc.; Hoboken, NJ, USA: 2013. Measurement of MHC/Peptide Interactions by Gel Filtration or Monoclonal Antibody Capture. Chapter 18Unit 18.3
23. Cheng Y, Prusoff WH. Relationship between the inhibition constant (K<sub>I</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochem. Pharmacol.* 1973; 22:3099–108. [PubMed: 4202581]
24. Gulukota K, Sidney J, Sette A, DeLisi C. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol.* 1997; 267:1258–1267. [PubMed: 9150410]
25. Yaciuk JC, Skaley M, Bardet W, Schafer F, Mojsilovic D, Cate S, Stewart CJ, McMurtrey C, Jackson KW, Buchli R, Olvera A, Cedeno S, Plana M, Mothe B, Brander C, West JT, Hildebrand WH. Direct Interrogation of Viral Peptides Presented by the Class I HLA of HIV-Infected T Cells. *J. Virol.* 2014; 88:12992–13004. [PubMed: 25165114]
26. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, Ly A, Lie W, Hildebrand WH, Mardis ER, Linette GP. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science.* 2015; 348:1–9.
27. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie G. a. Ma B. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol. Cell. Proteomics.* 2012; 11:M111.010587–M111.010587.
28. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Røder G, Peters B, Sette A, Lund O, Buus S. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One.* 2007; 2:e796. [PubMed: 17726526]

29. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*. 2009; 61:1–13. [PubMed: 19002680]
30. Jørgensen KW, Rasmussen M, Buus S, Nielsen M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology*. 2014; 141:18–26. [PubMed: 23927693]
31. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*. 2015; 43:D405–D412. [PubMed: 25300482]
32. Granados DP, Sriranganadane D, Daouda T, Zieger A, Laumont CM, Caron-Lizotte O, Boucher G, Hardy M-P, Gendron P, Côté C, Lemieux S, Thibault P, Perreault C. Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides. *Nat. Commun*. 2014; 5:3600. [PubMed: 24714562]
33. Marcilla M, Alpízar A, Lombardía M, Ramos-Fernandez A, Ramos M, Albar JP. Increased diversity of the HLA-B40 ligandome by the presentation of peptides phosphorylated at their main anchor residue. *Mol. Cell. Proteomics*. 2014; 13:462–74. [PubMed: 24366607]
34. Thommen DS, Schuster H, Keller M, Kapoor S, Weinzierl AO, Chennakesava CS, Wang X, Rohrer L, von Eckardstein A, Stevanovic S, Biedermann BC. Two preferentially expressed proteins protect vascular endothelial cells from an attack by peptide-specific CTL. *J. Immunol*. 2012; 188:5283–92. [PubMed: 22544926]
35. Kotturi MF, Peters B, Buendia-Laysa F, Sidney J, Oseroff C, Botten J, Grey H, Buchmeier MJ, Sette A. The CD8+ T-cell response to lymphocytic choriomeningitis virus involves the L antigen: uncovering new tricks for an old virus. *J. Virol*. 2007; 81:4928–4940. [PubMed: 17329346]
36. Wenzel T, Eckerskorn C, Lottspeich F, Baumeister W. Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. *FEBS Lett*. 1994; 349:205–209. [PubMed: 8050567]
37. Nussbaum AK, Dick TP, Keilholz W, Schirle M, Stevanovic S, Dietz K, Heinemeyer W, Groll M, Wolf DH, Huber R, Rammensee H-G, Schild H. Cleavage motifs of the yeast 20S proteasome subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci*. 1998; 95:12504–12509. [PubMed: 9770515]
38. Kisselev AF, Akopian TN, Woo KM, Goldberg AL. The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *J. Biol. Chem*. 1999; 274:3363–3371. [PubMed: 9920878]
39. van Endert PM, Tampé R, Meyer TH, Tisch R, Bach JF, McDevitt HO. A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity*. 1994; 1:491–500. [PubMed: 7895159]
40. Schumacher TN, V Kantesaria D, Heemels MT, Ashton-Rickardt PG, Shepherd JC, Fruh K, Yang Y, Peterson PA, Tonegawa S, Ploegh HL. Peptide length and sequence specificity of the mouse TAP1/TAP2 translocator. *J. Exp. Med*. 1994; 179:533–540. [PubMed: 8294864]
41. Chang S-C, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism. *Proc. Natl. Acad. Sci. U. S. A*. 2005; 102:17107–17112. [PubMed: 16286653]
42. Tynan FE, Borg NA, Miles JJ, Beddoe T, El-Hassen D, Silins SL, M. van Zuylen WJ, Purcell AW, Kjer-Nielsen L, McCluskey J, Burrows SR, Rossjohn J. High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T-cell receptor engagement and T-cell immunodominance. *J. Biol. Chem*. 2005; 280:23900–9. [PubMed: 15849183]
43. Burrows SR, Rossjohn J, McCluskey J. Have we cut ourselves too short in mapping CTL epitopes? *Trends Immunol*. 2006; 27:11–6. [PubMed: 16297661]
44. Samino Y, López D, Guil S, Saveanu L, van Endert PM, Del Val M. A long N-terminal-extended nested set of abundant and antigenic major histocompatibility complex class I natural ligands from HIV envelope protein. *J. Biol. Chem*. 2006; 281:6358–65. [PubMed: 16407287]
45. Hörig H, Young AC, Papadopoulos NJ, DiLorenzo TP, Nathenson SG. Binding of longer peptides to the H-2Kb heterodimer is restricted to peptides extended at their C terminus: refinement of the

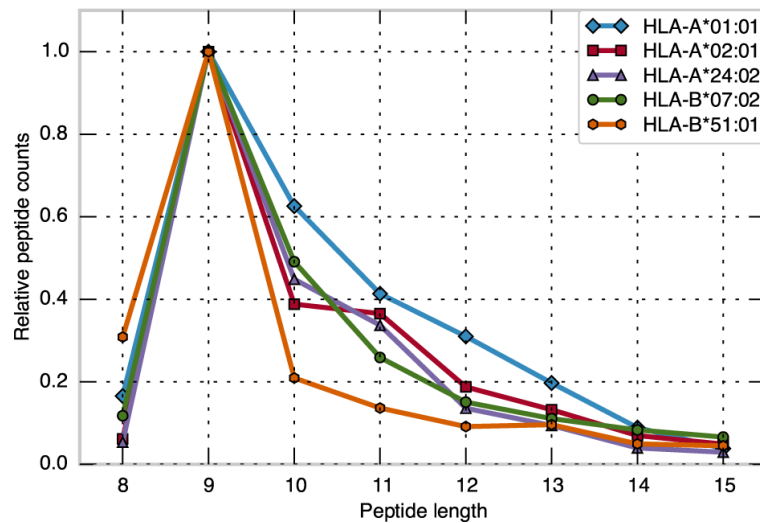
- inherent MHC class I peptide binding criteria. *J. Immunol.* 1999; 163:4434–41. [PubMed: 10510385]
46. Collins EJ, Garboczi DN, Wiley DC. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature.* 1994; 371:626–9. [PubMed: 7935798]





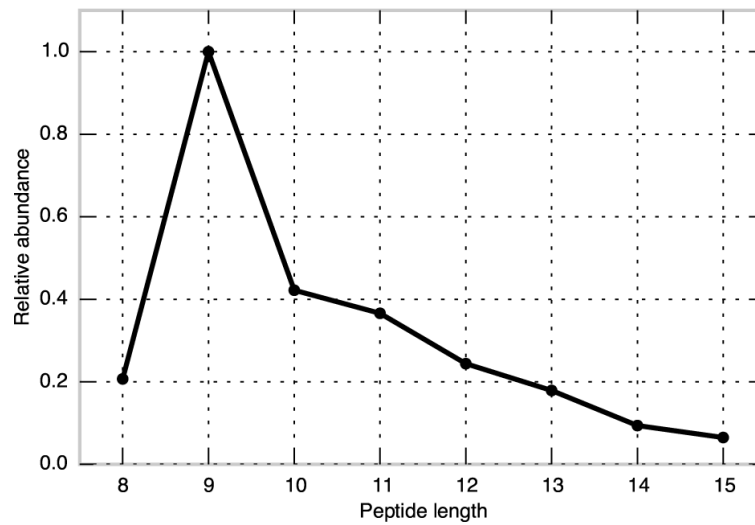
**Figure 1. Peptide binding length preference for five common HLA alleles**

The length preference for each HLA was determined by measuring the binding affinity of a series of fixed C-terminal combinatorial libraries of different length. Three series were tested, with either I, K or F at the C-terminal. The series with the strongest binding affinity was selected to represent the HLA allele. The selected series is denoted in the parentheses in the legend.  $IC_{50}$  binding affinities for each length were calculated as geometric means of 3-6 experiments. The relative binding affinities plotted were calculated as  $IC_{50}(9)/IC_{50}(L)$  where L is the peptide length. Error bars indicate standard errors of the geometric means.



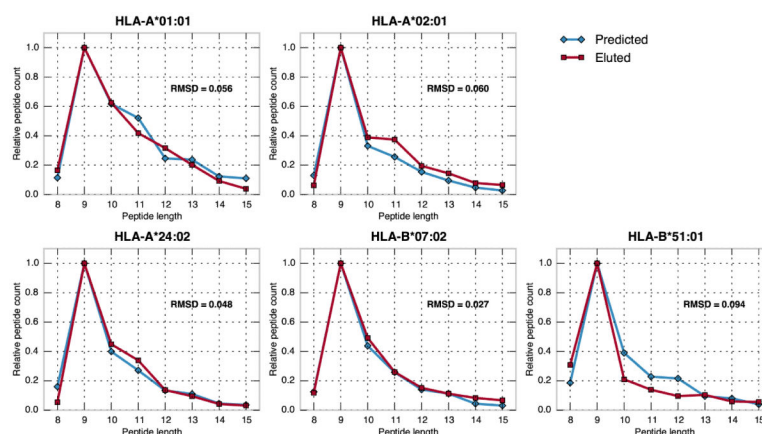
**Figure 2. Length profiles of naturally presented peptides for five HLA molecules**

Large datasets of HLA-I ligands were determined by the elution of ligands from secreted HLAs followed by mass spectrometry identification of the peptide sequences. From these ligand datasets, the number of ligands of each length was totaled. The y-axis indicates the number of ligands identified for a given length normalized by the number of peptides identified for the HLA at length 9.



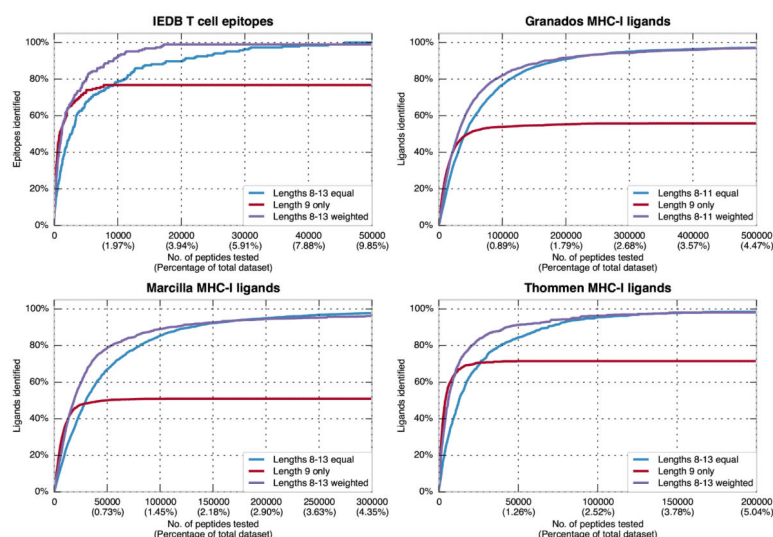
**Figure 3. Model fit of the available peptide length profile**

The available peptide length profile was fitted using MHC ligand length profiles and HLA binding length preferences for HLA-A\*01:01, HLA-A\*02:01, HLA-A\*24:02, HLA-B\*07:02 and HLA-B\*51:01 as described in the Materials and Methods. The optimal value for  $\beta$  associated with the fit was 0.30.



**Figure 4. Predicted vs. measured ligand length profiles for five HLA molecules**

A leave-one-out training was carried out by removing an HLA from the training dataset and then fitting the available peptide length profile with the remaining four HLAs. The resulting available peptide length profile was used in conjunction with the removed HLA's binding length preference (Fig. 1) to predict the removed HLA's ligand length profile. This predicted length profile was then compared to the measured ligand length profile of the removed HLA. As an example, in the HLA-A\*01:01 plot, HLA-A\*01:01 data was not used to fit an available peptide length profile (not shown). This available peptide length profile was then combined with the HLA-A\*01:01 binding length preference to determine the predicted ligand length profile (blue line). This profile was compared to the measured HLA-A\*01:01 ligand length profile (red line).



**Figure 5. Benchmarks of T cell epitope and MHC-I ligand predictions**

For each benchmark dataset, source proteins for each of the epitopes/ligands were downloaded and split into overlapping peptides of various lengths. The lengths of the overlapping peptides were determined by the lengths of the epitopes/ligands in the benchmark datasets; 8-13mer overlapping peptides for the IEDB, Marcilla and Thommen datasets, and 8-11mers for the Granados dataset. For each dataset, three sorted peptide lists were created using the following approaches: 1) predict affinities for all overlapping peptides and rank them based on their predicted  $IC_{50}$  value without taking length into account, 2) predict affinities for all 9mer peptides and rank them based on their predicted  $IC_{50}$  values (peptides of other lengths are considered non-candidates), 3) predict length corrected binding affinities for all overlapping peptides using the novel method described here and rank the peptides based on length corrected predictions. The plots show the number of epitopes/ligand identified by each approach as a function of the number of peptides tested, had the peptides been selected using the sorted lists described above.

**Table I**

Peptides identified in elution studies

Allele	Preferred peptide length for binding <sup>*</sup>	Fixed C-terminal library selected	No. of eluted ligands in final dataset	Most frequent eluted ligand length <sup>†</sup>
HLA-A*01:01	10	F	2992	9
HLA-A*02:01	9	I	4456	9
HLA-A*24:02	9	F	3949	9
HLA-B*07:02	9	F	3314	9
HLA-B*51:01	8	I	1275	9

<sup>\*</sup> The preferred peptide length identified from the MHC-I binding length preferences shown in Fig. 1.

<sup>†</sup> The preferred peptide length identified from the eluted MHC-I ligand profiles shown in Fig. 2.

**Table II**

Fitted ER peptide length profile.

Length	ER*
8	0.207
9	1.000
10	0.422
11	0.366
12	0.244
13	0.179
14	0.094
15	0.065

\* Fitted ER peptide length profile using data from the five HLA alleles, HLA-A\*01:01, HLA-A\*02:01, HLA-A\*24:02, HLA-B\*07:02 and HLA-B\*51:01. The  $\beta$  value associated with this profile was 0.3